



Analyst Perspective

David Menninger
SVP & Research
Director

Data Virtualization Brings Data Together Quickly and Easily

The technology industry throws around a lot of similar terms with different meanings as well as entirely different terms with similar meanings. In this post, I don't want to debate the meanings and origins of different terms; rather, I'd like to highlight a technology weapon that you should have in your data management arsenal. We currently refer to this technology as data virtualization. Other similar terms you may have heard include data fabric, data mesh and [data] federation. I'll briefly discuss these terms and how I see them being used, but ultimately, I'd like to share with you some research that shows why data virtualization can be valuable, regardless of what you call it.

We define data virtualization as the process of combining data on the fly from multiple sources rather than copying that data into a common repository such as a data warehouse or a data lake. Data virtualization requires the collection and presentation of metadata to make the collection of data appear as if it were one repository. It also requires intelligence in the process of bringing data together to enhance performance in the data retrieval process — for example, caching results of queries so subsequent similar queries do not require accessing the source systems.

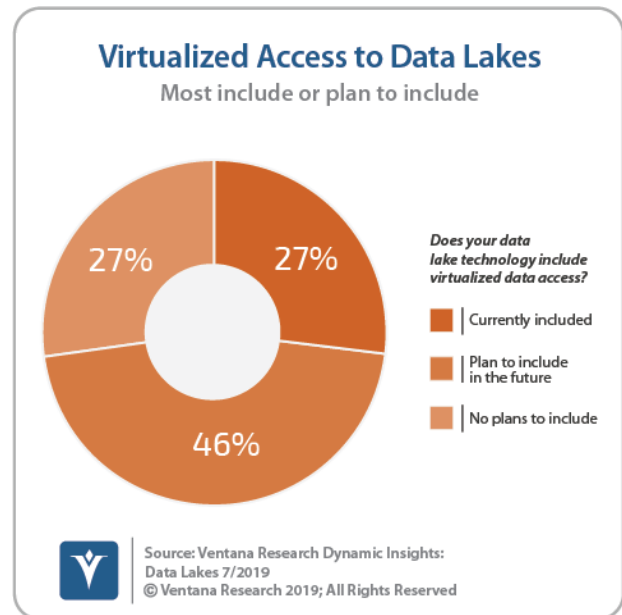
Data virtualization encompasses many more sophisticated techniques, such as evaluating whether to temporarily move some data from one system to another to execute a join. This approach could dramatically increase performance if the join involves a small table and a very large table. And now with artificial intelligence and machine learning, the evaluation of how to execute queries over distributed data sources can be quite sophisticated.

I prefer the term data virtualization over data fabric because I have seen the latter term used to describe data movement from its source to another repository, and perhaps also supporting some data virtualization. On the other hand, data mesh appears to refer less to the way in which data is brought together, and instead refers to the ownership of the data and metadata — the premise being that different sets of data and metadata should be managed and owned by different domain experts. Finally, [data] federation is a term



that generally refers to point-to-point connections for access to remote data sources such as external tables.

Our research shows that data virtualization is popular in the big data world. One-quarter (27%) of participants in our [Data Lake Dynamic Insights](#) research reported they are currently using data virtualization, and another two-quarters (46%) plan to include data virtualization in the future. Even more interesting, those who are using data virtualization reported higher rates of satisfaction (79%) with their data lake than those who are not (36%). Data virtualization makes it easier to incorporate a wide variety of data sources into a logical data lake. Data virtualization also helps organizations with data governance, since data remains in the source platform, ensuring all access rights are maintained in a single system. As a result, [data platform](#) vendors are including data virtualization capabilities in products.



Data virtualization is not a substitute for a data lake or data warehouse. It will never provide the same performance as consolidated data would provide. However, performance is not the only consideration. Given the proliferation of data sources and data sovereignty issues, it may be the only way to incorporate certain data sources into a single, logical repository. In some situations, eliminating the costs associated with storing and managing duplicated data may justify the performance trade-off implied in virtualized access to the data. In any case, organizations should have the option as part of an information architecture to utilize data virtualization where appropriate.

Regards,

David Menninger
SVP & Research Director

To read more perspectives by Dave, visit <https://davidmenninger.ventanaresearch.com/>



David Menninger – SVP and Research Director, Ventana Research

David Menninger is responsible for the overall research direction of data, information and analytics technologies at Ventana Research covering major areas including Analytics, Big Data, Business Intelligence and Information Management along with the additional specific research categories including Information Applications, IT Performance Management, Location Intelligence, Operational Intelligence and IoT, and Data Science.